

Η μετεξέλιξη ενός αποθετηρίου κοινωνικών επιστημών: το παράδειγμα της Πάνδημου σε περιβάλλον Fedora

Παρασκευάς Καμάτσος, Αφροδίτη Σωτηροπούλου, Άννα Θεοφιλάτου,
Ελευθερία Μπογιατζή, Κωνσταντία Κακάλη

Περίληψη

Η υλοποίηση ενός Ιδρυματικού Αποθετηρίου αποτελεί μία σύνθετη διαδικασία ανάλυσης των αναγκών του ακαδημαϊκού ιδρύματος που εξυπηρετεί, αξιολόγησης και επιλογής λογισμικού εφαρμογής, καθορισμού πολιτικών τεκμηρίωσης, αξιοποίησης μοντέλων διαλειτουργικότητας και μακροπρόθεσμης διατήρησης του περιεχομένου.

Στην περίπτωση της Βιβλιοθήκης του Παντείου Πανεπιστημίου, προτάθηκε ένα πλαίσιο ανάπτυξης της Ψηφιακής Βιβλιοθήκης «Πάνδημος», που βασίζεται στην επιλογή του λογισμικού Fedora για την εγκατάσταση και λειτουργία του. Η επιλογή του Fedora έγινε λόγω των προηγμένων δυνατοτήτων διαχείρισης και αναπαράστασης συλλογών και αντικειμένων, που σε συνδυασμό με την ευέλικτη αρχιτεκτονική του, επέτρεψε την ενσωμάτωση ετερογενούς ψηφιακού υλικού (γκρίζα βιβλιογραφία, φωτογραφικό και ιστορικό αρχείο, ψηφιοποιημένο υλικό συνεδρίων, περιοδικών κ.α.), την υπηρεσία ηλεκτρονικής κατάθεσης υλικού, τεκμηρίωσης και δημοσίευσής του, καθώς και την περαιτέρω βελτίωση και προβολή του περιεχομένου (SEO) απόδοσής του σε πολλαπλά σχήματα μεταδεδομένων κ.ά.

Στην παρούσα εισήγηση περιγράφονται τα στάδια επιλογής λογισμικού, διαχείρισης και οργάνωσης συλλογών και η υλοποίησή τους με ανάπτυξη χωριστών περιβαλλόντων εισαγωγής δεδομένων (editors). Επιπλέον αναλύονται τα πρότυπα διαλειτουργικότητας που εφαρμόζονται και ο καθορισμός πολιτικών περιγραφής του υλικού με την επιλογή σχημάτων μεταδεδομένων. Παράλληλα, επισημαίνονται τα προβλήματα που επιλύθηκαν σε κάθε στάδιο υλοποίησης του ιδρυματικού αποθετηρίου, τα βήματα που ακολουθήθηκαν για την αντιμετώπισή τους, καθώς και οι συνεργασίες που αναπτύχθηκαν με τα εμπλεκόμενα μέλη του προσωπικού της Βιβλιοθήκης (πληροφορικών & βιβλιοθηκονόμων) και ο βαθμός αλληλεξάρτησης των εργασιών τους.

Τέλος, παρουσιάζεται η ιδιαιτερότητα της ανάπτυξης μίας ψηφιακής βιβλιοθήκης σε λογισμικό Fedora σε ελληνικό ακαδημαϊκό ίδρυμα, η τεχνογνωσία που αποκτήθηκε και οι προοπτικές για το μέλλον.

Λέξεις Κλειδιά

Ιδρυματικά αποθετήρια, Ψηφιακές βιβλιοθήκες, Λογισμικό ανοιχτού κώδικα, Διαλειτουργικότητα, Σχήματα μεταδεδομένων, Institutional repositories, Digital libraries, Open source Software, Interoperability, Metadata schemas

1. Εισαγωγή

Η πορεία εξέλιξης ενός ιδρυματικού αποθετηρίου καθορίζεται από παράγοντες που σχετίζονται τόσο με τη δομή και τη λειτουργία του ιδρύματος που ανήκει (διάρθρωση τμημάτων και σχολών, προγράμματα σπουδών κ.ά.) όσο και από τις ευρύτερες εθνικές και διεθνείς πολιτικές και τάσεις στον ακαδημαϊκό χώρο και την επιστημονική κοινότητα (ανοικτή πρόσβαση, ακαδημαϊκές συνεργασίες, πολιτικές δημοσιεύσεις κ.α.). Η ψηφιακή βιβλιοθήκη “Πάνδημος” εκκινώντας ως ένα αποθετήριο κατάθεσης μεταπτυχιακών εργασιών και διδακτορικών διατριβών (institutional mandate) του Παντείου Πανεπιστημίου, σταδιακά μετεξελίχθηκε σε ένα ευρύτερο χώρο δημοσίευσης επιστημονικού υλικού των κοινωνικών επιστημών. Η πολιτική ενσωμάτωσης περιοδικών, συνεδρίων και δημοσιεύσεων οδήγησε στην αλλαγή της μορφής και του περιεχομένου, συμβάλλοντας στην δημιουργία ενός ψηφιακού αποθετηρίου που περιλαμβάνει:

- Γκρίζα βιβλιογραφία (διδακτορικά, μεταπτυχιακά, σημειώσεις μαθημάτων)
- Περιοδικά κοινωνικών επιστημών (Ουτοπία, Σύγχρονα Θέματα, Τετράδια κ.ά.)
- Συνέδρια της επιστημονικής κοινότητας
- Δημοσιεύσεις μελών Δ.Ε.Π. του Παντείου Πανεπιστημίου
- Αρχαιακό υλικό του Πανεπιστημίου

Οι παράγοντες που συνετέλεσαν στην ανάπτυξη και τον εμπλουτισμό του αποθετηρίου, προσδίδοντας τη σημερινή του μορφή, είναι από τη μία πλευρά η αξιοποίηση των έργων ψηφιοποίησης της Βιβλιοθήκης για την ένταξη των περιοδικών και των συνεδρίων στη συλλογή και από την άλλη το κύρος του ίδιου του Παντείου Πανεπιστημίου που λειτουργήσε θετικά στην εξασφάλιση των αδειών δημοσίευσης του υλικού από τους κατόχους των δικαιωμάτων.

Η συσσωμάτωση αυτού του υλικού στη συλλογή της “Πάνδημου”, πέρα από τις σημαντικές λειτουργίες της μακροπρόθεσμης διατήρησής της, της διάθεσής της προς όλους και της προβολής της, προσφέρει και ένα επιπλέον χαρακτηριστικό: την ενοποίηση της επιστημονικής και ακαδημαϊκής γνώσης στον ίδιο ψηφιακό χώρο, διευκολύνοντας την επικοινωνία μεταξύ των μελών της ακαδημαϊκής κοινότητας (Van de Sompel, Herbert et al., 2004).

2. Απαραίτητες λειτουργικότητες Συστήματος Ψηφιακής Βιβλιοθήκης (ΣΨΒ)

Οι συλλογές της “Πάνδημου” χαρακτηρίζονται από ευρεία *ανομοιογένεια* και ποικιλομορφία. Κάθε συλλογή έχει συγκεκριμένη επιστημονική ή πολιτιστική σημασία, περιλαμβάνει διάφορων τύπου υλικό (π.χ. κείμενο και χειρόγραφα, περιοδικά, φωτογραφίες, κ.α.), είτε αμιγώς ψηφιακό (born digital), είτε ψηφιοποιημένο (digitized), ικανοποιώντας διαφορετικές απαιτήσεις των χρηστών σε ό,τι αφορά τη δομή, τα μεταδεδομένα και την παρουσίαση.

Επιπρόσθετα, οι περισσότερες συλλογές χαρακτηρίζονται από *πολυπλοκότητα* στη δομή. Η κάθε συλλογή έχει τη δική της εσωτερική δομή και ιεραρχία, ενώ διαμορφώνονται διαφορετικές συσχετίσεις μεταξύ των αντικειμένων της (Pygounakis, G., Saidis, K., Nikolaidou, M., & Lourdi, I., 2004).

Η έρευνα ξεκίνησε με την απαίτηση για ένα ψηφιακό αποθετήριο με ευρεία ποικιλία χαρακτηριστικών, συμπεριλαμβανομένης της κλιμακοθετησιμότητας (scalability) για να χειριστεί εκατοντάδες εκατομμύρια ψηφιακών πόρων, της ευελιξίας για να χειριστεί τη συνεχώς διευρυνόμενη λίστα ψηφιακών μορφών και της επεκτασιμότητας για να διευκολυνθεί η ανάπτυξη προσαρμόσιμων εργαλείων και

υπηρεσιών, που μπορούν να συνεργάζονται με το αποθετήριο (Payette, S., & Staples, T., 2002).

Πιο συγκεκριμένα, ένα ΣΨΒ πρέπει να είναι *παραμετροποιήσιμο* σε μεγάλο βαθμό για να μπορεί να προσαρμοστεί στις ανάγκες της εκάστοτε συλλογής, υπάρχουσας ή μελλοντικής. Πρέπει να υποστηρίζει ψηφιακές συλλογές, που αποτελούνται από διαφορετικές κατηγορίες αντικειμένων, με ετερογενή και ανομοιόμορφα χαρακτηριστικά, παρέχοντας όμως μια ενιαία διεπαφή προς τους τελικούς χρήστες μέσα από τον παγκόσμιο ιστό.

Βασική απαίτηση είναι η υποστήριξη ανοικτών πρότυπων σε όλους τους τομείς, ώστε να εξασφαλίζεται η βιωσιμότητά του, η δυνατότητα αναβάθμισής του, η μελλοντική μετάπτωσή του σε άλλο σύστημα και η συνεργασία του με άλλες εφαρμογές λογισμικού ή ανάλογα συστήματα ψηφιακής βιβλιοθήκης. Αναγκαία είναι η εν δυνάμει υποστήριξη οποιοδήποτε πρότυπου μεταδεδομένων ή επεκτάσεις του καθώς και τοπικών συνόλων μεταδεδομένων απαραίτητων για μια συλλογή. Επιπλέον, είναι επιθυμητή η συμβατότητα του συστήματος με τα υπάρχοντα πρότυπα διαλειτουργικότητας μεταξύ διαφορετικών Ψηφιακών Βιβλιοθηκών, εφόσον υπάρχει η προοπτική συνεργασίας μιας Ψηφιακής Βιβλιοθήκης με άλλων Ακαδημαϊκών Ιδρυμάτων, όπως ήδη συμβαίνει και με ανάλογες ψηφιακές υπηρεσίες των βιβλιοθηκών. Γι' αυτό το λόγο, πρέπει να υποστηριχθεί το πρότυπο Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)¹ (Lagoze, C., & Van de Sompel, H., 2001), που επιτρέπει το διαμοιρασμό των μεταδεδομένων μεταξύ Ψηφιακών Βιβλιοθηκών. Τέλος, το σύστημα ψηφιακής βιβλιοθήκης θα πρέπει να είναι *ευέλικτο και επεκτάσιμο*, ώστε να επιτρέπει την προσθήκη επιπλέον λειτουργικότητας και την παραμετροποίησή του με προκαθορισμένο τρόπο.

3. Αναβάθμιση και κριτήρια επιλογής νέου ΣΨΒ

Σύμφωνα με τη λίστα συστημάτων Ψηφιακής Βιβλιοθήκης που προτείνει το Open Society Institute (Open Society Institute, 2004), για τον καθορισμό της νέας πλατφόρμας ανάπτυξης της “Πάνδημου”, εξετάστηκαν τα συστήματα ψηφιακών βιβλιοθηκών DSpace, Fedora, Eprints, Greenstone, Keystone και CDS (Kökörçený, M., & Bodnárová, A., 2010), (Pyrounakis, G., Nikolaidou, M., 2009), καταλήγοντας σε πρώτο στάδιο στα DSpace² και Fedora³, λόγω της στιβαρότητάς τους σε σύγκριση με τα υπόλοιπα (Εικόνα 1).

Η επιλογή του λογισμικού για το νέο ΣΨΒ “Πάνδημος”, με στόχο τη χρησιμοποίηση λογισμικού ανοικτού κώδικα (open source software) και βασισμένο σε JAVA τεχνολογίες, έγινε με τα εξής κριτήρια (Pyrounakis, G., Saidis, K., Nikolaidou, M., & Lourdi, I., 2004):

- Χαμηλό κόστος εγκατάστασης και συντήρησης.
- Επεκτασιμότητα και ευελιξία του συστήματος.
- Ευρεία χρήση από ακαδημαϊκά ιδρύματα.
- Πολλές δυνατότητες που προσφέρουν οι κοινότητες χρηστών και προγραμματιστών που υποστηρίζουν αυτά τα λογισμικά

¹ Open Source Initiative, <http://www.openarchives.org/>

² DSpace Federation, <http://www.dspace.org/>

³ Fedora Project, <http://www.fedora.info/>

	DSPACE	FEDORA
Πρότυπα και πρωτόκολλα - Μεταδεδομένα	Dublin Core	Οποιοδήποτε σχήμα
Πρότυπα και πρωτόκολλα - Αντικείμενα	Όλοι οι γνωστοί τύποι αρχείων. Κωδικοποίηση METS, OAI-PMH.	Όλοι οι γνωστοί τύποι αρχείων. Κωδικοποίηση FO-XML, METS, OAI-PMH.
Πλατφόρμες που υποστηρίζονται	Unix/ Linux	Unix/ Linux, Windows
Γλώσσα προγραμματισμού/ βάσεις δεδομένων/ εξυπηρετητής	Java/ PostgreSQL	Java, MySQL, PHP, Perl
Διεπαφή Διαχειριστή	Μέσω Διαδικτύου δημιουργία «κοινοτήτων», διαχείριση ομάδων χρηστών, μεταδεδομένων	Μέσω αυτόνομης εφαρμογής, προσφέρονται όλες οι διαδικασίες δημιουργίας και αλληλεπίδρασης με το Απόθεμα.
Διεπαφή Χρήστη	Πλοήγηση, Αναζήτηση σε οποιοδήποτε πεδίο, πλήρες κείμενο. Προσωπικός χώρος, προφίλ.	Προσφέρεται μόνο αναζήτηση στο Απόθεμα. Επιπλέον υπηρεσίες πρέπει να αναπτυχθούν.
Πολύγλωσσία	Πολύγλωσσες διεπαφές και μεταδεδομένα με μικρή τροποποίηση.	Όχι στις ήδη ανεπτυγμένες διεπαφές. Ναι στο Απόθεμα.
Επεκτασιμότητα	Μέτρια - Plugin manager, μηχανισμός τροποποίησης, επιπρόσθετοι μηχανισμοί, επέκταση πλαισίου	Άριστη - SOAP, Web Service Interface
Αποθήκευση (Storage)	Backups, Auditing, Distribution, DBMS + Bitstreams	XML Storage (FOXML)
Οριμότητα/Υποστήριξη	Πολύ Καλή	Πολύ Καλή
Επιπρόσθετα Χαρακτηριστικά	Στατιστικά, LDAP.	Διατήρηση Ιστορικού και εκδόσεων Αντικειμένων. Έτοιμα εργαλεία από τρίτους.

Εικόνα 1 – Σύγκριση DSpace και Fedora

3.1. Fedora και βασικά κριτήρια επιλογής του

Με βάση τα χαρακτηριστικά των συλλογών της “Πάνδημου” και τις απαραίτητες λειτουργικότητες που πρέπει να πληροί μια ψηφιακή βιβλιοθήκη οριοθετήθηκαν οι απαιτήσεις και τέθηκαν οι βασικές προδιαγραφές του ΣΨΒ: το σχήμα μεταδεδομένων που θα χρησιμοποιεί, τις διάφορες πολιτικές πρόσβασης - διαχείρισης που θα υλοποιεί, τα θέματα καθιερωμένων όρων, οι δυνατότητες προσβασιμότητας-αναζήτησης, οι δυνατότητες συμβατότητας με άλλα συστήματα. Περαιτέρω, λαμβάνοντας υπόψη τη *συγκριτική μελέτη λογισμικών*, επιλέχθηκε το Fedora για τους παρακάτω λόγους (Pyrounakis,G., Saidis,K., Nikolaidou,M., & Lourdi, 2004) (Kökörcsený, M., & Bodnároná, A., 2010):

- Ευέλικτη αρχιτεκτονική
- Υποστήριξη ετερογενών μορφών δεδομένων και ενσωμάτωσή τους στην περίπτωση που εμφανισθούν νέες μορφές
- Υποστήριξη της ανάπτυξης εξωτερικών modules που επικοινωνούν εύκολα με το σύστημα μέσω των προσφερόμενων διεπαφών (APIs)
- Ομαδοποίηση διαφορετικών δεδομένων σε σύνθετα αντικείμενα
- Παρουσίαση των αντικειμένων με πολλές διαφορετικές «όψεις» του περιεχομένου
- Διαχωρισμός των δεδομένων μεταξύ των διεπαφών χρήσης και των μηχανισμών παρουσίασης των δεδομένων
- Βασική οντότητα το ψηφιακό αντικείμενο που αποθηκεύεται σαν αρχείο XML (εμπεριέχει μοναδικό ID, μεταδεδομένα, datastreams και disseminators). Οι disseminators επεκτείνουν τις «όψεις» των ψηφιακών αντικειμένων συνδέοντας

το περιεχόμενό τους με λογισμικό (εργαλεία και/ή υπηρεσίες) που παρέχουν τη δυνατότητα μορφοποίησης του περιεχομένου με τέτοιο τρόπο, έτσι ώστε να ταιριάζει με την «άποψη» του χρήστη, ή να μπορεί ακόμη να χρησιμοποιηθεί από κάποιο λογισμικό.

➤ Μεγάλες δυνατότητες επέκτασης γιατί:

- Περιλαμβάνει ευρετήρια όρων (βασισμένα στα μεταδεδομένα)
- Βασίζεται η ευρετηρίαση των μεταδεδομένων στο πρότυπο DC και παράλληλα υλοποιείται μέσω εξωτερικών υπηρεσιών
- Πραγματοποιούνται συσχετίσεις μεταξύ διαφορετικών συλλογών
- Γίνεται λεπτομερής αναζήτηση πεδίων (field search) στο πλαίσιο κάθε συλλογής χωριστά.

Το Fedora βασίζεται στην αρχιτεκτονική Flexible Extensible Digital Object and Repository Architecture (FEDORA). Η ανάπτυξή του ξεκίνησε το 2001 από τα πανεπιστήμια University of Virginia και Cornell University. Η υλοποίηση του λογισμικού έχει πραγματοποιηθεί σε γλώσσα Java και είναι διαθέσιμο σε πλατφόρμες Windows και Unix/Linux, ενώ χρησιμοποιεί τις τεχνολογίες του Ιστού (Web Technologies) προκειμένου να επιτευχθεί η μέγιστη δυνατή διαλειτουργικότητα και επεκτασιμότητα στις ψηφιακές συλλογές.

Το Fedora έχει δύο βασικές οντότητες. Το *ψηφιακό αντικείμενο* (digital object) και το *αποθετήριο* (repository). Το ψηφιακό αντικείμενο είναι βασισμένο στη δομή που ορίζει το πρότυπο Metadata Encoding and Transmission Standard (METS)⁴. Αποτελείται από ένα μοναδικό προσδιοριστή (PID), τα μεταδεδομένα συστήματος, ένα ή περισσότερα datastream (που αποτελούν τα μεταδεδομένα ή το ψηφιακό υλικό) και τα disseminator που συνδέουν τα datastream με κάποια συμπεριφορά. Υπάρχουν τριών ειδών ψηφιακά αντικείμενα: τα αντικείμενα δεδομένων (data objects), τα αντικείμενα καθορισμού συμπεριφοράς (behavior definition objects) και τα αντικείμενα μηχανισμού συμπεριφοράς (behavior mechanism objects). Τα αντικείμενα δεδομένων αποτελούν τις οντότητες που περιέχουν το ψηφιακό υλικό και τα μεταδεδομένα, ενώ τα υπόλοιπα δύο καθορίζουν και υλοποιούν τις μεθόδους για την παρουσίαση ή τη μετατροπή των datastream ενός αντικειμένου δεδομένων.

Το Fedora παρουσιάζει την πιο ευέλικτη αρχιτεκτονική, ενώ το μοντέλο ψηφιακών αντικειμένων, που υποστηρίζει, προσφέρει τεράστιες δυνατότητες επέκτασης. Αυτός είναι και ο σημαντικότερος λόγος επιλογής του, όπως έχει ήδη αναφερθεί. Πιο συγκεκριμένα, οι περισσότερες επεκτάσεις βασίζονται στα datastreams, από τα οποία απαρτίζεται κάθε ψηφιακό αντικείμενο. Έτσι, παρέχεται η δυνατότητα υποστήριξης σύνθετων αντικειμένων καθώς και η ενσωμάτωση ετερογενών και ποικίλων μεταδεδομένων (περιγραφικά, τεχνικά, διαχειριστικά ή δομικά). Επιπλέον, με τη χρήση συμπεριφορών στα ψηφιακά αντικείμενα είναι δυνατό να υλοποιηθούν συσχετίσεις μεταξύ αντικειμένων με δυναμικό και ενιαίο τρόπο. Υποστηρίζεται, επίσης, η δυνατότητα χρήσης προτύπων (templates) για τη δημιουργία, επεξεργασία και παρουσίαση των ψηφιακών αντικειμένων. Τέλος, οι διαφορετικές συλλογές και υποσυλλογές μπορούν να αναπαρασταθούν ως κανονικά ψηφιακά αντικείμενα, προσφέροντας ένα ενιαίο και κομψό σχήμα αναπαράστασης στο προγραμματιστικό επίπεδο.

Η αρχιτεκτονική του συστήματος βασίζεται στο διαχωρισμό των διαθέσιμων υπηρεσιών σε τέσσερα υποσυστήματα: υποσύστημα αποθήκευσης (storage subsystem), υποσύστημα διαχείρισης (management subsystem), υποσύστημα

⁴ Metadata Encoding and Transmission Standard (2004), <http://www.loc.gov/standards/mets/>

ασφάλειας (security subsystem) και υποσύστημα πρόσβασης (access subsystem). Παρέχει τη δυνατότητα ανάπτυξης πρόσθετης λειτουργικότητας, η οποία ούτε περιορίζει, αλλά ούτε και περιορίζεται από την υπάρχουσα διάρθρωση του συστήματος, καθώς κάθε «υποσύστημα» μπορεί να λειτουργεί ξεχωριστά και απομονωμένα από τα υπόλοιπα, διατηρώντας όμως την ικανότητα επικοινωνίας μέσω του βασικού Fedora API. Έτσι, επιτυγχάνεται η ανάπτυξη ενός ΣΨΒ «πάνω από» το Fedora, το οποίο μπορεί να λειτουργεί ανεξάρτητα σε ό,τι αφορά την επιχειρησιακή του λογική, τα μεταδεδομένα που χρησιμοποιεί και τη σημασία που αυτά έχουν, παραμένοντας συμβατό με τις νέες εκδόσεις του Fedora. Κατά συνέπεια, οι χρήστες επιδρούν με το περιεχόμενο του αποθετηρίου μέσω client εφαρμογών, φυλλομετρητών, προγραμμάτων ή εφαρμογών που εκτελούνται σε κάποιο εξυπηρετητή (server applications) (Payette, S., & Staples, T., 2002) (Payette, S., & Lagoze, C., 1998).

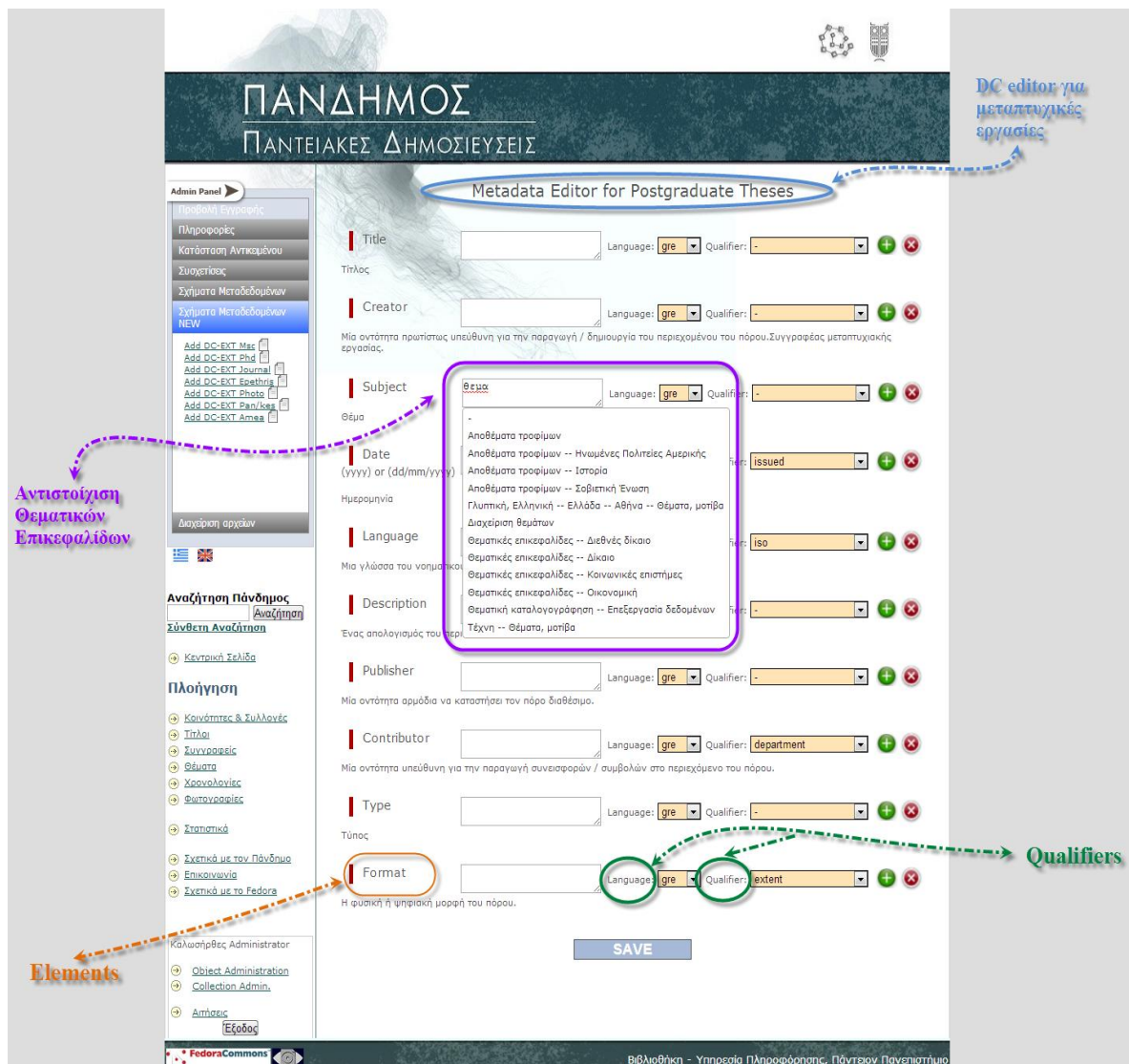
3.2. Μετάπτωση δεδομένων από παλιό σύστημα στο νέο

Σε συνέχεια της υλοποίησης και παραμετροποίησης του νέου ΣΨΒ, πραγματοποιήθηκε η μετάπτωση των δεδομένων από το υπάρχον σύστημα στο νέο. Αυτή η διαδικασία ήταν ιδιαίτερα κρίσιμη καθώς υπάρχει πάντα ο κίνδυνος να χαθούν δεδομένα. Μια από τις βασικές απαιτήσεις ήταν και η διατήρηση των τεκμηρίων που ήδη περιείχε το ψηφιακό αποθετήριο, καθώς η μεταφορά του περιεχομένου του αποθετηρίου σε ένα καινούριο σύστημα συνεπάγεται διατήρηση της σύνδεσης μεταξύ μεταδεδομένων και τεκμηρίων, διατήρηση της δομής των τελευταίων, αλλά και των σχέσεων που συνδέουν τα τεκμήρια μεταξύ τους. Το μοντέλο οργάνωσης της πληροφορίας διαφέρει εν μέρει από το ένα σύστημα στο άλλο, λόγω και της διαφορετικότητας του λογισμικού, για αυτό δόθηκε ιδιαίτερη προσοχή στο πως θα αντιστοιχηθούν σωστά οι έννοιες των κοινοτήτων – συλλογών του DSpace στις συλλογές – υποσυλλογές του νέου συστήματος (Fedora). Τα υπάρχοντα 2200 περίπου ψηφιακά αντικείμενα από το DSpace αποθετήριο, μετανάστευσαν στο νέο ΣΨΒ “Πάνδημο”, χωρίς να χάσουν κανένα δεδομένο πληροφορίας, στη πλατφόρμα του Fedora. (Gavrilis, D., Aggelis St. & Papatheodorou C., 2011).

3.3. Dublin Core Editors

Για την υλοποίηση του Dublin Core Editor στο ψηφιακό αποθετήριο, απαραίτητες ήταν οι οδηγίες για την τεκμηρίωση των διαφορετικών τύπου συλλογών, σύμφωνα με τη μελέτη των βιβλιοθηκονόμων, οι οποίες αφορούν ουσιαστικά την περιγραφή των μεταδεδομένων που γίνεται με χρήση του προτύπου Dublin Core, ενώ για την υλοποίηση του DC editor χρησιμοποιήθηκαν οι τεχνολογίες: PHP, HTML, MySQL και JavaScript. Συγκεκριμένα, σε κάθε Dublin Core Editor παρουσιάζονται αναλυτικά τα elements και τα qualifiers που χρησιμοποιούνται για την περιγραφή του υλικού. Όσο πιο πλούσια είναι αυτά (άρα και η εγγραφή ενός χρήστη), τόσο πιο "προσπελάσιμο" θα είναι ένα τεκμήριο (πχ. Εικόνα 2).

Ο DC editor χρησιμοποιείται για να υποβάλλονται τεκμήρια σε μια συλλογή. Οι πληροφορίες που συμπληρώνονται στη φόρμα θα διαμορφώσουν την εγγραφή των μεταδεδομένων, δίνοντας τη δυνατότητα στους χρήστες να ανακτούν το τεκμήριο τους μέσω των μηχανών αναζήτησης.



Εικόνα 2 – Dublin Core Editor για μεταπτυχιακές εργασίες

Για την ακόμα καλύτερη τεκμηρίωση των συλλογών, υλοποιήθηκαν οκτώ εκδόσεις του editor, λόγω μικρών διαφορών και πιο εξειδικευμένων elements και qualifiers σε κάθε είδους τεκμηρίου : 1) Μεταπτυχιακές εργασίες, 2) Διδακτορικές διατριβές, 3) Άρθρα περιοδικών, 4) Πανεπιστημιακές εκδόσεις, 5) Επετηρίδες, 6) Υλικό για Αμεα, 7) Φωτογραφίες / Εικόνες και 8) Συλλογές -Υποσυλλογές.

3.4. Θεματική ευρετηρίαση με ελεγχόμενο λεξιλόγιο

Ένας βασικός στόχος της ανάπτυξης του νέου ΣΨΒ “Πάνδημος” ήταν ο έλεγχος των καθιερωμένων όρων και η ακριβής συσχέτισή τους με το αρχείο καθιερωμένων όρων που αναπτύσσει παράλληλα η βιβλιοθήκη για το συμβατικό περιβάλλον (Γαβρίλης Δ., Κακάλη, Κ & Διακάκη Α., 2007) Έτσι, χρησιμοποιούνται οι θεματικές επικεφαλίδες στα ελληνικά και στα αγγλικά με την εισαγωγή τους από εφαρμογή POLARIS, που έχει αναπτυχθεί από τη Βιβλιοθήκη για την διαχείριση των εγγραφών καθιερωμένων όρων (<http://library.panteion.gr/polaris>), όπως δείχνει η Εικόνα παραπάνω. Τα θέματα, στο πεδίο Subject των DC Editors, κατά την καταχώρηση από τους βιβλιοθηκονόμους, αντλούνται αυτόματα από το Polaris, μια βάση όλων των

θεματικών επικεφαλίδων του Παντείου, ώστε να υπάρχει απόλυτη ταύτιση με τα ήδη καταχωρημένα θέματα. Έτσι πληκτρολογώντας τουλάχιστον δύο χαρακτήρες, είτε ελληνικούς είτε λατινικούς, με δυνατότητες auto-complete, εμφανίζεται η λίστα θεμάτων που περιλαμβάνει τους χαρακτήρες αυτούς, διευκολύνοντας και τυποποιώντας τη διαδικασία καταγραφής του πεδίου, με αποτέλεσμα να μπορούν να παρέχονται αξιόπιστες υπηρεσίες πλοήγησης και αναζήτησης στους χρήστες με βάση το θέμα. Επιπλέον, με την λειτουργία της υπηρεσίας αυτοκατάθεσης προβάλλονται οι λέξεις – κλειδιά των συγγραφέων.

4. Μετάβαση από το αποθετήριο στη ψηφιακή βιβλιοθήκη.

Οι δυνατότητες του Fedora ως προς την παραμετροποίηση της αναζήτησης των ψηφιακών πόρων και τη διασυνδεσιμότητα μεταξύ των συλλογών, αλλά και ο αρχικός στόχος, που είχε τεθεί, άνοιξαν το δρόμο προς τη μετάβαση από ένα ακαδημαϊκό αποθετήριο σε μια ψηφιακή βιβλιοθήκη, που θεματικά εναρμονίζεται με το επιστημονικό πεδίο του πανεπιστημίου, παρέχοντας μια πιο ολοκληρωμένη παροχή υπηρεσιών προς το χρήστη. Στις συλλογές που αφορούσαν δημοσιεύσεις της συγγραφικής δραστηριότητας των μελών του πανεπιστημίου, δηλ. μεταπτυχιακές εργασίες, διδακτορικές διατριβές και σημειώσεις μαθημάτων προστέθηκε μεταξύ άλλων και το ιστορικό και φωτογραφικό αρχείο του πανεπιστημίου.

4.1 Το φωτογραφικό αρχείο του Παντείου Πανεπιστημίου

Η συλλογή του φωτογραφικού αρχείου του Πανεπιστημίου αφορά υλικό που τοποθετείται από τη δεκαετία του 1950 έως και σήμερα. Πρόκειται για φωτογραφίες που θεματικά καλύπτουν εκδηλώσεις που πραγματοποιήθηκαν, επετείους, αναγορεύσεις επίτιμων διδασκόντων, εγκαίνια κτιρίων καθώς επίσης και επισκέψεις επισήμων προσώπων στη Σχολή.

Κύριο χαρακτηριστικό της συλλογής είναι ότι έχει μορφή χρονολογίου. Οι φωτογραφίες αποτυπώνουν γεγονότα που έλαβαν χώρα κατά τη διάρκεια της λειτουργίας του πανεπιστημίου και θεωρήθηκαν σημαντικά για να καταγραφούν. Η φωτογραφική συλλογή επομένως, αποτελείται κυρίως από ιστορικές φωτογραφίες συμπληρώνοντας έτσι το ιστορικό αρχείο του ιδρύματος. Με την ένταξη του υλικού αυτού στον Πάνδημο επιτυγχάνεται πρωτίστως η διατήρησή του και κατ' επέκταση η ανάδειξη, η δημοσιοποίηση και προβολή της ιστορίας του πανεπιστημίου, καθώς επίσης και η διάθεσή του για ακαδημαϊκή και ιστορική έρευνα. Επιπλέον, μελλοντικός στόχος είναι και ο συσχετισμός του με τις υπόλοιπες συλλογές του αποθετηρίου ενισχύοντας τη δυνατότητα των διασυνδεόμενων πόρων.

Το φυσικό αρχείο, το οποίο φυλασσόταν έως τώρα στην αίθουσα του Πρυτανικού Συμβουλίου, αποτελείται από έγχρωμες και ασπρόμαυρες φωτογραφίες οργανωμένες σε φωτογραφικά άλμπουμ αλλά και σε φακέλους, χωρίς συγκεκριμένη ταξινόμηση. Το έργο της ψηφιοποίησης του υλικού αυτού ξεκίνησε πιλοτικά και παραδόθηκε μέρος του φωτογραφικού αρχείου στη Βιβλιοθήκη το 2011. Ο φωτογράφος που ανέλαβε τη ψηφιοποίηση παρέδωσε σύμφωνα με τις προδιαγραφές που είχαν καθοριστεί, 1.000 περίπου φωτογραφίες, οργανωμένες σε folder που αντιστοιχούν στα φωτογραφικά άλμπουμ/φακέλους και εσωτερικά σε υποφακέλους αρχείων τριών διαφορετικών μεγεθών, σε μορφή tiff (για το αρχείο) και jpeg (medium και thumbnail για τη ψηφιακή βιβλιοθήκη). Η μη συστηματική ταξινόμηση, η ελλιπής τεκμηρίωση μεγάλου μέρους του αρχείου, καθώς επίσης και η δομή

αρχειοθέτησής του, η οποία διατηρήθηκε κατά τη ψηφιοποίηση, αποτέλεσαν καθοριστικούς παράγοντες για το σχεδιασμό της δομής της συλλογής αυτής.

4.2 Η επεξεργασία και ένταξη του φωτογραφικού αρχείου στον Πάνδημο

Η διαχείριση της πρόσκτησης αποτέλεσε μια χρονοβόρα διαδικασία προετοιμασίας του ψηφιακού αρχείου για την ένταξή του στη ψηφιακή βιβλιοθήκη. Αυτό περιλαμβάνει τον έλεγχο ποιότητας, την ταυτοποίηση με το φυσικό αρχείο, την καταγραφή και κωδικοποίηση πληροφοριών που προκύπτουν από το φυσικό αρχείο (αρίθμηση, τοποθεσία, τίτλος, χρονολογία, μέγεθος, φωτογράφος κτλ).

Ακολούθησε ο καθορισμός της δομής των συλλογών στον Πάνδημο και η επιλογή σχήματος μεταδεδομένων και ελεγχόμενων λεξιλογίων για την περιγραφική και θεματική καταλογογράφηση. Η δομή των συλλογών και υποσυλλογών κάτω από τις οποίες εντασσονται τα ψηφιακά αντικείμενα αποφασίστηκε σύμφωνα με τα όσα αναφέρθηκαν να έχει την εξής δενδροειδή μορφή :

- ◆ Φωτογραφικό Αρχείο Παντείου Πανεπιστημίου
 - ◆ 2007-2011 Πρυτανεία Παναγιώτη Τσίρη
 - ◆ Εκδήλωση για τα 80 χρόνια Πάντειο Πανεπιστήμιο
 - ◆ 80 χρόνια Πάντειο στιγμιότυπο 1

Η ένταξη των φωτογραφιών τοποθετείται σε πρώτο επίπεδο χρονολογικά, δηλ. ανά πρυτανεία και σε δεύτερο, θεματικά, ανά εκδήλωση-γεγονός. Οι υποσυλλογές που ανήκουν κάτω από κάθε πρυτανεία και η καταλογογράφησή τους προέκυψαν ύστερα από μια διαδικασία μελέτης του κάθε φακέλου, εντοπισμού και αναγνώρισης προσώπων, κυρίως των πρυτάνεων, για μια πρώτη χρονολογική τοποθέτηση και οργάνωσης καταλόγων ονομάτων και βιογραφικών. Στη μελέτη αυτή χρησιμοποιήθηκαν στοιχεία από τις επετηρίδες του Παντείου Πανεπιστημίου, από πηγές στο ίντερνετ (Wikipedia, τέως ψηφιακό αρχείο της ΕΡΤ, online εφημερίδες, ιστοσελίδα του Παντείου Πανεπιστημίου) καθώς επίσης και από συνεντεύξεις με συναδέλφους και καθηγητές στο Πάντειο.

Αντίστοιχα, για το καθορισμό της φόρμας μεταδεδομένων ορίστηκαν τα βασικά κριτήρια για την περιγραφή των φωτογραφιών, παράγοντες που ορίζουν τα χαρακτηριστικά της συλλογής, όπως ο τύπος του πόρου, η φύση του και οι ανάγκες των χρηστών (Jung-ran Park and Yuji Tosaka, 2010). Ελήφθησαν, επίσης, υπόψη βασικές θεωρίες για την κατανόηση και ευρετηρίαση των εικόνων. Για παράδειγμα το μοντέλο που πρότεινε η Sara Shatford (1986), βασιζόμενη στη θεωρία του ιστορικού τέχνης Erwin Panofsky, που διακρίνει την περιγραφή των εικόνων σε τρία επίπεδα: *Generic* (pre-iconographic), *Specific* (iconographic) and *Abstract* (iconological) αναλύοντας το καθένα από αυτά σε τέσσερις όψεις *Who*, *What*, *Where* and *When*.

Στο πλαίσιο της δυνατότητας χρήσης πολλαπλών προτύπων μεταδεδομένων ακολούθησε μια βασική επισκόπηση προτύπων είτε γενικών, όπως το UNIMARC και το Dublin Core αλλά και ειδικών. Συγκεκριμένα, εξετάστηκε το VRA 4.0. Core⁵, πρότυπο για την περιγραφή εικόνων, έργων τέχνης και πολιτισμού, το CCO⁶, πρότυπο για την περιγραφή πολιτιστικών αντικειμένων, καθώς και το IPTC Photo Metadata Standard⁷, πρότυπο για ψηφιακές κάμερες και λογισμικά εικόνας.

⁵ VRA Core 4.0 <http://www.vraweb.org/projects/vracore4/index.html>

⁶ CCO (Cataloguing Cultural Objects) <http://www.vraweb.org/ccoweb/cco/about.html>

⁷ IPTC Standard Photo Metadata http://www.iptc.org/std/photometadata/documentation/IPTC-CS5-FileInfo-UserGuide_6.pdf

Το σχήμα μεταδεδομένων που τελικά αποφασίστηκε να εφαρμοστεί για να περιγράψει τις συλλογές και υποσυλλογές του φωτογραφικού αρχείου είναι το Dublin Core, συμβατό με το πρωτόκολλο OAI-PMH metadata harvesting. Η χρήση του προτύπου είναι ευρέως διαδεδομένη και χρησιμοποιείται σε υψηλό ποσοστό (Dublin Core qualified 40,6% και unqualified 25,4%) από ψηφιακά αποθετήρια σε σχέση με εξειδικευμένα πρότυπα (π.χ. VRA 14,9%) (Jung-ran Park & Yuji Tosaka, 2010).

Επιπροσθέτως, καθώς η “Πάνδημος” συγκεντρώνει και διαχειρίζεται πολλές και διαφορετικές κατηγορίες υλικού με επικρατέστερο τύπο ψηφιακών πόρων το κείμενο, όπως εξάλλου και στα περισσότερα αποθετήρια (text 86,5%), η χρήση εικόνων αποτελεί το 1/7 του συνόλου της συλλογής, των οποίων οι ανάγκες, τα χαρακτηριστικά και το μέγεθος που πρόκειται να φιλοξενηθούν καλύπτονται από τις δυνατότητες του DC.

Για τη θεματική ευρετηρίαση, αντίστοιχα, αποφασίστηκε να χρησιμοποιηθούν οι θεματικές επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου, επικρατέστερη επιλογή σε αποθετήρια που χειρίζονται εικόνες (LCSH 43,2%), καθώς επίσης και ο θησαυρός όρων της Βιβλιοθήκης του Κογκρέσου για γραφικό υλικό, TGMII⁸. Για τις ανάγκες του φωτογραφικού αρχείου μεταφράστηκαν πιθανοί όροι με σκοπό να χρησιμοποιηθούν ως θησαυρός όρων που λειτουργούν ως genre headings και προσδιορίζουν το είδος της εικόνας (πχ. Πορτρέτα, Ομαδικά πορτρέτα, Φωτογραφία τοπίου, Αρχιτεκτονική φωτογραφία κτλ).

5. Αυτοκατάθεση Μεταπτυχιακών Εργασιών και Διδακτορικών Διατριβών

Αυτοκατάθεση είναι ένας ευρύς όρος, που χρησιμοποιείται, όταν ο συγγραφέας υποβάλλει χωρίς την μεσολάβηση τρίτου την ηλεκτρονική μορφή της εργασίας του σε μία ψηφιακή βιβλιοθήκη ή αποθετήριο. (Xia, J., & Sun, L., 2007). Η γρήγορη πρόσβαση στην εργασία των άλλων που προσφέρει η “Πάνδημος” είναι πολύ χρήσιμο εργαλείο για την έρευνα, αλλά τα βασικά οφέλη για τους μεταπτυχιακούς φοιτητές και υποψήφιους διδάκτορες σχετίζονται με την προώθηση της δικής τους εργασίας. Οι φοιτητές στην ουσία είναι οι δημιουργοί του αποθετηρίου και οι χρήστες του. Η ευκαιρία που τους δίνεται να παρουσιάσουν τα ευρήματά τους σε ηλεκτρονική μορφή, τους επιτρέπει να εκφραστούν με πιο ευέλικτους και δημιουργικούς τρόπους. (Copeland, S., Penman, A., & Milne, R., 2005).

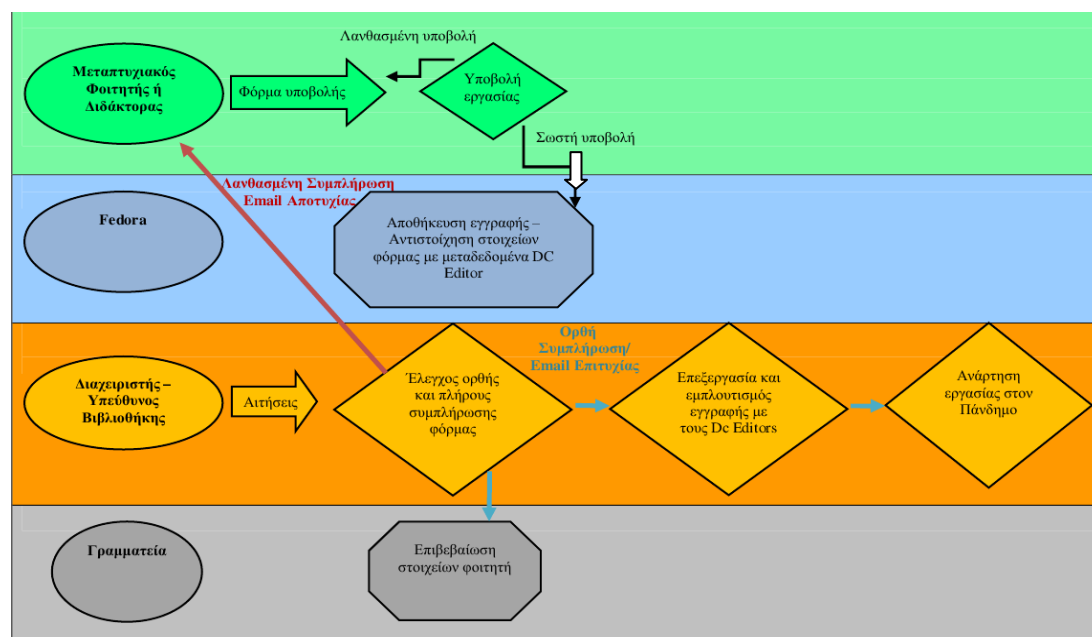
Οι μεταπτυχιακές εργασίες και οι διδακτορικές διατριβές αντανakλούν την ικανότητα του ιδρύματος να καθοδηγεί τους φοιτητές και να υποστηρίζει την πρωτότυπη έρευνα. Καθώς το αποθετήριο του Πανεπιστημίου γίνεται πιο δημοφιλές, οι φοιτητές και το ακαδημαϊκό προσωπικό θα προβούν σε κρίσεις σχετικά με την ποιότητα του πανεπιστημίου μέσω της ψηφιακής του βιβλιοθήκης. Στο μέλλον, η ποιότητα ενός πανεπιστημίου θα συνδέεται με το αποθετήριό του. (Moxley, J. M., 2001).

Η Βιβλιοθήκη του Παντείου Πανεπιστημίου σε συνεργασία με τη διοίκηση του Πανεπιστημίου έκανε από 1/4/2013 υποχρεωτική την αυτοκατάθεση μεταπτυχιακών εργασιών και διδακτορικών διατριβών στην “Πάνδημος” από τους μεταπτυχιακούς φοιτητές και διδάκτορες αντίστοιχα. Η διαδικασία της αυτοκατάθεσης έχει γίνει πολύ απλή χάρη σε μια ξεχωριστή διαδικτυακή (*web based*) φόρμα υποβολής, η οποία δρα ως ενδιάμεσος μεταξύ του Fedora και του τελικού χρήστη. Η κατάθεση των

⁸ TGM II (Thesaurus for Graphic Material II <http://www.loc.gov/rr/print/tgm2/>)

τεκμηρίων γίνεται μέσω της φόρμας υποβολής εργασίας, η οποία διατίθεται στην κεντρική σελίδα της “Πάνδημου”, ενώ η διεπαφή της φόρμας και η σχετική βοήθεια είναι σε δύο γλώσσες, ελληνικά και αγγλικά.

Η φόρμα περιλαμβάνει όλα τα απαραίτητα πεδία για την πλήρη περιγραφή του υλικού από τον φοιτητή. Έχουν δημιουργηθεί δύο φόρμες υποβολής, μία για τους μεταπτυχιακούς φοιτητές και μια για τους διδάκτορες. Όλα τα πεδία της φόρμας είναι υποχρεωτικά. Αφού γίνει η ηλεκτρονική κατάθεση της εργασίας, γίνεται έλεγχος από το προσωπικό, στέλνεται με ηλεκτρονικό μήνυμα η βεβαίωση κατάθεσης, η οποία είναι απαραίτητη για την υποβολή αίτησης ορκωμοσίας στη γραμματεία του τμήματος.



Εικόνα 3 - Ροή εργασιών αυτοκατάθεσης

Αφού ολοκληρωθεί η πρώτη φάση της κατάθεσης από τον φοιτητή, ακολουθεί άμεσα η επεξεργασία από το προσωπικό της Βιβλιοθήκης. Δίνεται η δυνατότητα απευθείας επεξεργασίας των στοιχείων της κατάθεσης στους editors της “Πάνδημου”, όπου κάθε τιμή που δίνει ο χρήστης σε κάθε πεδίο της φόρμας χρησιμοποιείται για την απόδοση τιμής στο αντίστοιχο πεδίο του σχήματος μεταδεδομένων Dublin Core στον editor, με το προσωπικό να χρειάζεται να κάνει τη θεματική επεξεργασία της εργασίας, την εισαγωγή του pdf αρχείου και τέλος την εισαγωγή των συσχετίσεων και την ενεργοποίηση της εγγραφής.

6. Ζητήματα προσβασιμότητας και ευρετηρίασης στις μηχανές αναζήτησης

Πέρα από την προφανή αξία της ευρετηρίασης των τεκμηρίων και της προβολής τους στο web, με την ταυτόχρονη πρόσβαση των χρηστών του διαδικτύου στην ακαδημαϊκή γνώση του πανεπιστημίου μέσω των ενεργειών που περιγράφηκαν, προσφέρονται επιπλέον δυνατότητες και εργαλεία για την μέτρηση της επιστημονικής απήχησης του έργου αυτού στην κοινότητα.

Ειδικότερα στον τομέα των παραδοσιακών βιβλιομετρικών μεθόδων η ένταξη της “Πάνδημου” στο Google Scholar, παρέχει την δυνατότητα παρακολούθησης των παραπομπών που καταγράφονται στην διεθνή βιβλιογραφία, για κάθε τεκμήριο της

ψηφιακής βιβλιοθήκης. Η πληροφορία αυτή είναι ιδιαίτερα σημαντική για την μέτρηση του κύρους των επιστημονικών δημοσιεύσεων του ιδρύματος και της κατάταξης του στις διεθνείς αξιολογήσεις.

Παράλληλα, η χρήση των Google Analytics προσφέρει την δυνατότητα καταγραφής εξίσου σημαντικών στατιστικών δεικτών, όπως είναι ο αριθμός των views κάθε εγγραφής και ο αριθμός των downloads των ψηφιακών τους αρχείων. Οι δείκτες αυτοί αποτελούν τμήμα των εναλλακτικών μεθόδων μέτρησης της ποιότητας των επιστημονικών έργων, που συγκροτούν τον όρο altmetrics (Priem et al., 2010).

Οι μηχανές αναζήτησης χρησιμοποιούν μεθόδους ευρετηρίασης και συμπερίληψης των ιστοσελίδων στα αποτελέσματα τους που βασίζονται σε crawlers. Μέσω αυτής της διαδικασίας οι ιστοσελίδες αποθηκεύονται σε διαφορετικούς χρόνους και αφού αναλύεται η δομή των δεδομένων και το περιεχόμενο, συγκρίνονται τα δύο «στιγμιότυπα» για τυχόν ενημερώσεις και προβάλλονται στα αποτελέσματα της αναζήτησης (Henzinger, Motwani, & Silverstein, 2002).

Αν και αυτή η μέθοδος λειτουργεί ικανοποιητικά για τις ιστοσελίδες γενικού περιεχομένου, δεν ανταποκρίνεται στην πραγματικότητα των μεταδεδομένων των βιβλιοθηκών και των ιδρυματικών αποθετηρίων, καθώς οι κατάλογοι τους βασίζονται σε παραδοσιακά βιβλιογραφικά δεδομένα όπως τίτλος, συγγραφέας και έτος έκδοσης (Lossau, 2004). Οι εγγραφές των τεκμηρίων μίας ψηφιακής βιβλιοθήκης συγκροτούνται από ένα σύνολο βιβλιογραφικών δεδομένων (τίτλος, συγγραφέας κ.α.), αρχείων (pdf, doc κ.α.) και ευρετηρίων πολλαπλών κριτηρίων (χρονολογικά, τίτλων, συγγραφέων, συλλογών κ.α.). Η ταυτοποίηση και ανάκληση των στοιχείων μίας βιβλιογραφικής εγγραφής παρουσιάζει προβλήματα στην ακρίβεια και την κατάταξη των αποτελεσμάτων ενός ερωτήματος. Η αύξηση της «ορατότητας» των τεκμηρίων ενός ιδρυματικού αποθετηρίου στα αποτελέσματα των μηχανών αναζήτησης, αποτελεί μία σύνθετη διαδικασία ενεργειών και ρυθμίσεων τόσο της δομής της ιστοσελίδας του αποθετηρίου όσο και της απόδοσης των μεταδεδομένων.

6.1. Πρωτοβουλίες για την «ορατότητα» και «προσβασιμότητα» στις μηχανές αναζήτησης

Παρατηρώντας τα αποτελέσματα στην αναζήτηση των τεκμηρίων στην Google, διαπιστώθηκε χαμηλή κατάταξη τους και σε ορισμένες περιπτώσεις απουσία ορισμένων τεκμηρίων από τα αποτελέσματα, ακόμη και αν ως όρος αναζήτησης είχε δοθεί ο τίτλος ή ο συγγραφέας του έργου. Παράλληλα, τον Μάιο του 2008 η Google ανακοίνωσε πως εγκαταλείπει την υποστήριξη του Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), το οποίο είχε συμπεριλάβει στα sitemaps των ιστοσελίδων που αναγνωρίζει και ευρετηριάζει, κρίνοντας πως είναι ασύμβατο με το πλήθος της πληροφορίας των ιδρυματικών αποθετηρίων που καλείται να διαχειριστεί (Google, 2008). Για την αντιμετώπιση αυτών των προβλημάτων ακολουθήθηκε μία σειρά αλλαγών στον χαρακτηρισμό των μεταδεδομένων των εγγραφών και της ιστοσελίδας της Ψηφιακής Βιβλιοθήκης. Οι κυριότερες ενέργειες :

i) Χρήση meta tags: Τα βιβλιογραφικά δεδομένα των τεκμηρίων, τροποποιήθηκαν ώστε να γίνουν «κατανοητά» από τις μηχανές αναζήτησης. Κάθε στοιχείο περιγραφής του σχήματος μεταδεδομένων Dublin Core, συντάχθηκε και ως meta tag. Το meta tag είναι ένα πολύ σημαντικό μέρος του κώδικα HTML της ιστοσελίδας μας, το οποίο δεν εμφανίζεται ως μέρος του περιεχομένου της ιστοσελίδας αλλά είναι αναγνωρίσιμο και παρέχει συγκεκριμένες πληροφορίες σχετικά με τη σελίδα σε

προγράμματα περιήγησης (ή άλλες υπηρεσίες web). Με την αλλαγή αυτή η εμφάνιση των δεδομένων στην html παρουσιάζει την μορφή :

```
<meta name='DC.creator' content='Κεφαλά, Άννα Χ.' />
<meta name='DC.date' content='2012' />
<meta name='DC.identifier' content='http://pandemos.panteion.gr/index.php?
op=record&type=0&q=&page=1&scope=0&lang=en&pid=iid:5396' scheme='DCTERMS.URI' />
<meta name='DC.description' content='Διατριβή (διδακτορική) - Πάντειο Πανεπιστήμιο. Τμήμα
Κοινωνιολογίας, 2012Βιβλιογραφία: σ. 282-306' xml:lang='en' />
<meta name='DC.subject' content='Urban ecology (Sociology)UrbanizationΑστική οικολογία
(Κοινωνιολογία)ΑστικοποίησηUrban ecology (Sociology)Urbanization' />
<meta name='DC.publisher' content='Πάντειο Πανεπιστήμιο Κοινωνικών και Πολιτικών Επιστημών' />
```

ii) Google webmaster tools: Έγινε εξέταση του site για προβλήματα στην ευρετηρίαση των σελίδων του (URLs) και για τον αριθμό των εγγραφών που εντοπίζονται. Ελέγχθηκε η ύπαρξη περιορισμών (robots.txt) που εμποδίζουν την πρόσβαση των μηχανών αναζήτησης στις σελίδες του site.

iii) Google analytics: Πραγματοποιήθηκε δήλωση του site της ψηφιακής βιβλιοθήκης και έγιναν οι κατάλληλες ρυθμίσεις.

Μετά από αυτές τις ενέργειες η παρουσία των τεκμηρίων στα αποτελέσματα των μηχανών αναζήτησης βελτιώθηκε σημαντικά, επιτυγχάνοντας τον στόχο της εμφάνισης των εγγραφών της “Πάνδημος” ως πρώτες στη σειρά κατάταξης. Παράλληλα ευρετηριάστηκε το σύνολο των URLs της ψηφιακής βιβλιοθήκης, αναγνωρίζοντας τα τεκμήρια, τις συλλογές και τα ευρετήρια των θεμάτων, συγγραφέων και των τίτλων.

6.2. Το ευρετήριο επιστημονικής βιβλιογραφίας Google Scholar και «Πάνδημος»

Η μηχανή αναζήτησης Google Scholar είναι η μεγαλύτερη, ελεύθερης πρόσβασης βάση ευρετηρίασης ακαδημαϊκής βιβλιογραφίας στο web. Περιλαμβάνει εκατομμύρια τεκμήρια από άρθρα περιοδικών, πρακτικά συνεδρίων, μονογραφίες, μελέτες, διατριβές κ.ά. Η μέθοδος αναζήτησης και συμπερίληψης επιστημονικών τεκμηρίων, είναι προσανατολισμένη στους εμπορικούς διαθέτες επιστημονικής πληροφόρησης, με κύριο ενδιαφέρον στα άρθρα περιοδικών. Ακολουθεί τα πρότυπα μεταδεδομένων που έχουν αναπτυχθεί από εκδότες όπως οι Atypon, Highwire Press, MetaPress. Αντίθετα τα ιδρυματικά αποθετήρια χρησιμοποιούν, στην πλειονότητα τους, το σχήμα μεταδεδομένων Dublin Core, που για το Google Scholar αποτελεί την τελευταία επιλογή απόδοσης βιβλιογραφικών δεδομένων (Google Scholar Inclusion Guidelines). Η διαφοροποίηση αυτή, ανάμεσα στα πρότυπα λειτουργίας των ιδρυματικών αποθετηρίων και τις προδιαγραφές του Google Scholar, προκαλεί δυσκολίες στην αναγνώριση των αποθετηρίων ως διαθέτες περιεχομένου επιστημονικής βιβλιογραφίας, με αποτέλεσμα τον αποκλεισμό τους.

Με δεδομένη την ασυμβατότητα των προτύπων του ιδρυματικού αποθετηρίου (DCMI, OAI) και του Google Scholar, έγινε προσπάθεια να εξεταστούν οι λόγοι για τον αποκλεισμό του αποθετηρίου και την έλλειψη παρουσίας των τεκμηρίων του στα αποτελέσματα του. Μελετήθηκαν αντίστοιχες περιπτώσεις αποθετηρίων από την διεθνή βιβλιογραφία και οι πιθανοί τρόποι αντιμετώπισης του προβλήματος. Τα περισσότερα ιδρύματα που το περιεχόμενο τους ευρετηριάζεται στο Google Scholar, χρησιμοποιούν λογισμικά αποθετηρίων που συγκαταλέγονται στα προτεινόμενα και συμβατά από το ίδιο το Scholar (DSpace, EPrints κ.α.). Η “Πάνδημος”, χρησιμοποιώντας το λογισμικό Fedora, δεν είχε την δυνατότητα να αναγνωριστεί το περιεχόμενο της «αυτόματα» χωρίς τις κατάλληλες τροποποιήσεις που το Google Scholar έχει θέσει ως προϋπόθεση για την συμπερίληψη τεκμηρίων

στα ευρετήρια του. Η έρευνα κατέληξε στην ανάγκη προσαρμογής των μεταδεδομένων στις απαιτήσεις του Google Scholar και σε μία σειρά ενεργειών παρουσίασης των εγγραφών και των συνοδευτικών ψηφιακών αρχείων.

Η διαδικασία ανάλυσης των δεδομένων του προβλήματος και οι προσπάθειες για τις απαιτούμενες αλλαγές, εξελίχθηκε σταδιακά και σε διάστημα τριών (3) περίπου μηνών, λόγω του απαιτούμενου χρόνου αναγνώρισης και ευρετηρίασης του περιεχομένου της ψηφιακής βιβλιοθήκης από το Google Scholar, που κυμαίνεται από 4-6 εβδομάδες (Google, 2013). Σύμφωνα με τις οδηγίες που παρέχονται από το Scholar, προτείνεται η χρήση των Highwire Press tags, EPrints tags ή των Be Press tags, και όχι η χρήση του Dublin Core, το οποίο προτείνεται ως τελευταία λύση. Λαμβάνοντας υπόψη τα παραπάνω και αντίστοιχες ενέργειες για το ίδιο πρόβλημα, άλλων ιδρυματικών αποθετηρίων (Arlitsch & O'Brien, 2012), προχωρήσαμε στις εξής βελτιώσεις:

i) Αντιστοιχήθηκαν τα elements του DC με τα Highwire Press tags: Πραγματοποιήθηκε σύνδεση των δύο σχημάτων, ώστε τα δεδομένα που εισάγονται να αποδίδονται ταυτόχρονα και στις δύο μορφές.

Dublin Core	Highwire Press tags
DC.title	citation_title
DC.creator	citation_author
DC.publisher	citation_publisher
DC.issued	citation_publication_date
DC.subject	citation_keywords
DC.description.abstract	citation_abstract_html_url
DC.contributor	citation_advisor
DC.contributor	citation_committee_member
DC.language	citation_language

ii) Browse by year: Δημιουργήθηκε λίστα με τα τεκμήρια της ψηφιακής βιβλιοθήκης κατά χρονολογική σειρά.

iii) Scholar Inclusion: Έγινε υποβολή του ιδρυματικού αποθετηρίου «Πάνδημος» στο Google Scholar, ώστε να αναγνωριστεί το περιεχόμενο του και να ευρετηριαστεί.

Μετά από τα παραπάνω, τα τεκμήρια του ιδρυματικού αποθετηρίου ευρετηριάστηκαν και συμπεριλήφθηκαν στο Google Scholar σε ποσοστό πολύ ικανοποιητικό (88.9/100), αν λάβει κανείς υπόψη του την απουσία ευρετηρίασης του πριν τις προαναφερθείσες ενέργειες. Συγκεκριμένα στα αποτελέσματα αναζήτησης του όρου: “site:pandemos.panteion.gr” ανακαλούνται 4250 αποτελέσματα/τεκμήρια της Ψηφιακής Βιβλιοθήκης «Πάνδημος», τα οποία είναι πλήρως αναζητήσιμα με πολλαπλά κριτήρια αναζήτησης (π.χ. τίτλος, συγγραφέας, έτος, κ.α.).

7. Συμπεράσματα και προοπτικές

Το ψηφιοποιημένο υλικό τεκμηριώνεται, δημιουργούνται μεταδεδομένα και συγκροτούνται συλλογές για την πληρέστερη αναπαράσταση των θεματικών ενοτήτων του. Για την τεκμηρίωση του υλικού χρησιμοποιείται η ίδια προσέγγιση με αυτήν της εσωτερικής επιστημονικής παραγωγής του πανεπιστημίου (LCSH, λέξεις κλειδιά, Dublin Core), συμβάλλοντας στην ομοιογένεια των μεταδεδομένων του αποθετηρίου, στην ευχερέστερη αναζήτηση και ανάκτηση του και στην διαλειτουργικότητα του με άλλα συστήματα. Οι χρήστες του αποθετηρίου έχουν την δυνατότητα της αναζήτησης, συλλογής, μελέτης, ανάλυσης δεδομένων και

δημιουργίας νέων έργων⁹ (Green, Ann G., 2007) από μία πλούσια συλλογή επιστημονικού υλικού, που αποτυπώνει την πορεία της ελληνικής σκέψης στον χώρο των κοινωνικών επιστημών τις τελευταίες δεκαετίες.

Εκτός από τις παραπάνω λειτουργικότητες της νεας ψηφιακής βιβλιοθήκης, έχουν σχεδιασθεί να αναπτυχθούν ακόμα: i) αυτόματη εξαγωγή πλήρους κειμένου με τεχνικές OCR για καλύτερη ευρετηρίαση και προηγμένες υπηρεσίες αναζήτησης και ii) δυνατότητα προσωποποιημένων υπηρεσιών προς τους χρήστες, που να μπορούν να εισάγουν και να διαχειρίζονται τα δικά τους δεδομένα. Στο πλαίσιο αυτό η “Πάνδημος” θα διασυνδεθεί με την εφαρμογή του σημασιολογικού εμπλουτισμού που έχει σχεδιαστεί και υλοποιείται (ονομάζεται **DLsem**) ως ένα ολοκληρωμένο σύστημα διεπαφής χρήστη με το σύνολο των πηγών της βιβλιοθήκης, καθώς και με τις επιμέρους υπηρεσίες της βιβλιοθήκης του Παντείου Πανεπιστημίου.

Σταθμίζοντας τα χαρακτηριστικά του αποθετηρίου, ως προς το περιεχόμενο και τους επιστημονικούς τομείς που καλύπτει, αλλά και των λειτουργιών και στόχων του εντός και εκτός του ιδρύματος, μπορεί να υποστηριχτεί η συγκρότηση του ως ένα θεματικό αποθετήριο (discipline repository) κοινωνικών επιστημών.

8. Βιβλιογραφία

- Arlitsch, K., & O'Brien, P. S. (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar. *Library Hi Tech*, 30(1), 60-81.
- Copeland, S., Penman, A., & Milne, R. (2005). Electronic theses: the turning point. *Program: Electronic library and information systems*, 39(3), 185-197.
- Curt, F. (2000, September 27). <http://computer.howstuffworks.com>. Ανάκτηση 4 23, 2013, από <http://computer.howstuffworks.com/internet/basics/search-engine.htm>
- Gavrilis D., Angelis S., Papatheodorou C. (2010). Mopseus – A Digital Repository System with Semantically Enhanced Preservation Services. In *International Conference on Preservation of Digital Objects (IPRES2010)*, Vienna Austria, September 2010, Proceedings, pp. 135-143, 2010.
- Green, A. G., & Gutmann, M. P. (2007). Building partnerships among social science researchers, institution-based repositories and domain specific data archives. *OCLC Systems & Services*, 23(1), 35-53.
- Google. (2008, Μάιος 23). Google Webmaster Central Blog. Ανάκτηση Απρίλιος 13, 2013, από <http://googlewebmastercentral.blogspot.gr/2008/04/retiring-support-for-oai-pmh-in.html>
- Google. (2013). Inclusion Guidelines for Webmasters. Ανάκτηση Μάρτιος 4, 2013, από <http://scholar.google.com/intl/en/scholar/inclusion.html#troubleshooting>
- Henzinger, M. R., Motwani, R., & Silverstein, C. (2002, September). Challenges in web search engines. In *ACM SIGIR Forum* (Vol. 36, No. 2, pp. 11-22). ACM.
- Kökörcseny, M., & Bodnárová, A. (2010, November). Comparison of digital libraries systems. In *Proceedings of the 9th WSEAS international conference on Data networks, communications, computers*, Faro, Portugal (pp. 97-100).
- Lagoze, C., & Van de Sompel, H. (2001, January). The Open Archives Initiative: Building a low-barrier interoperability framework. In *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries* (pp. 54-62). ACM.

- Lossau, N. (2004). Search Engine Technology and Digital Libraries: libraries Need to Discover the Academic Internet. Spring Forum meeting (pp. 12). New Orleans: American Digital Library Federation (DLF).
- Moxley, J. M. (2001). Universities should require electronic theses and dissertations. *Educause Quarterly*, 24(3), 61-63.
- Open Society Institute, A Guide to Institutional Repository Software. 2nd Edition, January 2004
- Park, J. R., & Tosaka, Y. (2013). Metadata creation practices in digital repositories and collections: Schemata, selection criteria, and interoperability. *Information Technology and Libraries*, 29(3), 104-116.
- Payette, S., & Lagoze, C. (1998). Flexible and extensible digital object and repository architecture (FEDORA). In *Research and Advanced Technology for Digital Libraries* (pp. 41-59). Springer Berlin Heidelberg.
- Payette, S., & Staples, T. (2002). The Mellon Fedora Project Digital Library Architecture Meets XML and Web Services. In *Research and Advanced Technology for Digital Libraries* (pp. 406-421). Springer Berlin Heidelberg.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto.
- Pyrounakis, G., & Nikolaidou, M. (2009). Comparing Open Source Digital Library Software.
- Pyrounakis, G., Saidis, K., Nikolaidou, M., & Lourdi, I. (2004). Designing an integrated digital library framework to support multiple heterogeneous collections. In *Research and Advanced Technology for Digital Libraries* (pp. 26-37). Springer Berlin Heidelberg.
- Shatford, S. (1986). Analyzing the subject of a picture: a theoretical approach. *Cataloging & classification quarterly*, 6(3), 39-62.
- Stodden, V. C. (2010). Data sharing in social science repositories: facilitating reproducible computational research
- Van de Sompel, H., Payette, S., Erickson, J., Lagoze, C., & Warner, S. (2004). Rethinking scholarly communication. *D-Lib Magazine*, 10(9), 1082-9873.
- Xia, J., & Sun, L. (2007). Assessment of self-archiving in institutional repositories: depositorship and full-text availability. *Serials Review*, 33(1), 14-21.
- Γαβρίλης, Δημήτρης., Κακάλη, Κ., & Διακάκη, Α. (2007). Θεματική ενοποίηση δημοσίων καταλόγων και ψηφιακών βιβλιοθηκών. Πρακτικά 16^{ου} Πανελληνίου Συνεδρίου ακαδημαϊκών Βιβλιοθηκών. Πειραιάς, Πανεπιστήμιο Πειραιώς 1-3 Οκτώβρη 2007.